UCRL-PRES-226277

Earth System Grid: Model Data Distribution & Server-Side Analysis to Enable Intercomparison Projects

PCMDI Software Team

Challenges facing ESG-CET

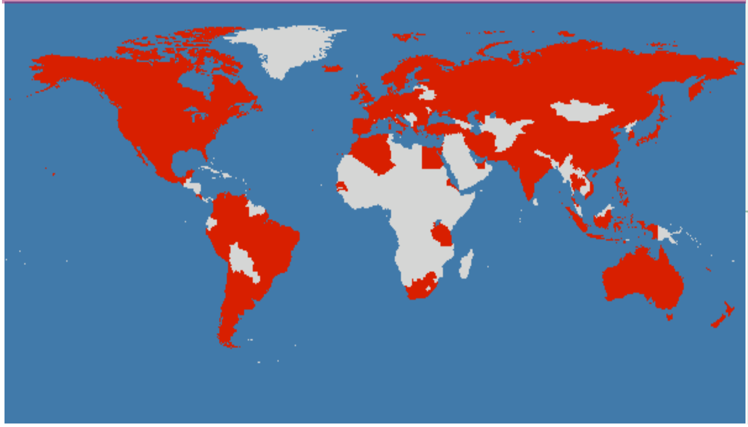
- Building on the very successful CMIP3 IPCC AR4 ESG data portal.
- How best to collect and distribute data on a much larger scale?
  - At each stage tools could be developed to improve efficiency
  - Substantially more ambitious community modeling projects (>~300 TBs) will require a distributed database
- Metadata describing extended modeling simulations (e.g., atmospheric aerosols and chemistry, carbon cycle, dynamic vegetation, etc.)
- How to make information understandable to end-users so that they can interpret the data correctly
- More users from WGI. (Possibly WGII and WGIII?)
- Client and Server-side analysis and visualization tools in a distributed environment (i.e., subsetting, concatenating, regridding, filtering, …)
- Testbed needed by late 2008 – early 2009

ESG facts and figures

ESG Objective

To support the infrastructural needs of the national and international climate community, ESG is providing crucial technology to securely access, monitor, catalog, transport, and distribute data in today's Grid computing environment.

Worldwide ESG user base



CMIP3 IPCC AR4 ESG Portal

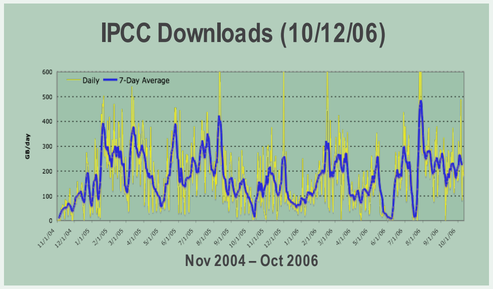
28 TB of data at the PCMDI site location

- 68,400 files
- Generated by a modeling campaign coordinated by the Intergovernmental Panel on Climate Change
- Model data from 11 countries

818 registered users

Downloads to date

- 123 TB
- 543,500 files
- 300 GB/day (average)



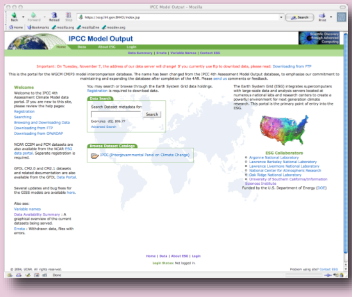
200 scientific papers published to date based on analysis of CMIP3 IPCC AR4 data

Providing climate scientists with virtual proximity to large simulation results needed for their research

ESG Goal

- Very large distributed data archives
  - ➢Easy federation of sites
  - ➢Across the US and around the world
- “Virtual Datasets” created through subsetting and aggregation
- Metadata-based search and discovery
- Web-based and analysis tool access
- Increased flexibility and robustness
- Server-side analysis

http://www-pcmdi.llnl.gov



IPCC Model Output – Mozilla

https://esg.llnl.gov:8443/index.jsp

IPCC Model Output

Home | Data | About ESG | Login

Data Summary | Errata | Variable Names | Contact ESG

Important: On Tuesday, November 7, the address of our data server will change! If you currently use ftp to download data, please read: Downloading from FTP

This is the portal for the WGCM CMIP3 model intercomparison database. The name has been changed from the IPCC 4th Assessment Model Output database, to emphasize our commitment to maintaining and expanding the database after completion of the 4AR. Please send us comments or feedback.

Welcome

Welcome to the IPCC 4th Assessment Climate Model data portal. If you are new to this site, please review the help pages:

Registration
Searching
Browsing and Downloading Data
Downloading from FTP
Downloading from OPeNDAP

NCAR CCSM and PCM datasets are also available from the NCAR ESG data portal. Separate registration is required.

GFDL CM2.0 and CM2.1 datasets and related documentation are also available from the GFDL Data Portal.

Several updates and bug fixes for the GISS models are available here.

Also see:

Variable names

Data Availability Summary : A graphical overview of the current datasets being served.

Errata : Withdrawn data, files with errors.

You may search or browse through the Earth System Grid data holdings. Registration is required to download data.

Data Search

Search Dataset metadata for: Search

Examples: c02, 806.77

Advanced Search

Browse Dataset Catalogs

IPCC (Intergovernmental Panel on Climate Change)

The Earth System Grid (ESG) integrates supercomputers with large-scale data and analysis servers located at numerous national labs and research centers to create a powerful environment for next generation climate research. This portal is the primary point of entry into the ESG.

ESG Collaborators

- Argonne National Laboratory
- Lawrence Berkeley National Laboratory
- Lawrence Livermore National Laboratory
- National Center for Atmospheric Research
- Oak Ridge National Laboratory
- University of Southern California/Information Sciences Institute

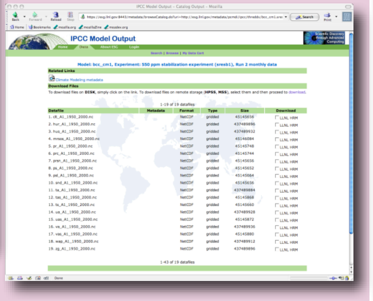
Funded by the U.S. Department of Energy (DOE)

Home | Data | About ESG | Login

Login Status: Not logged in.

© 2004, UCAR. All rights reserved.

Problem using site? Contact ESG



IPCC Model Output – Catalog Output – Mozilla

https://esg.llnl.gov:8443/metadata/browseCatalog.do?uri=http://esg.llnl.gov/metadata/pcmdi/ipcc/thredds/bcc_cm1.sres

IPCC Model Output

Home | Data | About ESG | Login

Search | Browse | My Data Cart

Model: bcc_cm1, Experiment: 550 ppm stabilization experiment (sresb1), Run 2 monthly data

Related Links

Climate Modeling metadata

Download Files

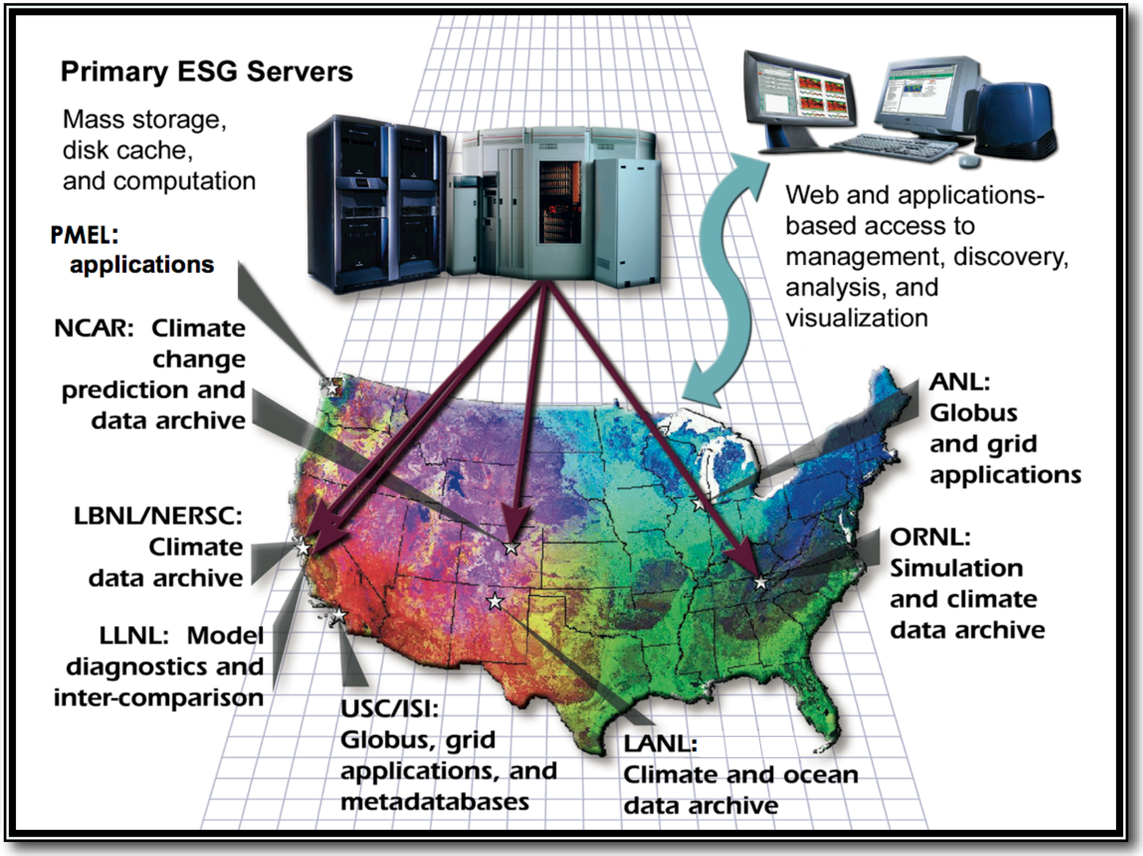
To download files on **DISK**, simply click on the link. To download files on remote storage (**HPSS**, **MSS**), select them and then proceed to download.

1-19 of 19 datafiles

Datafile	Metadata	Format	Type	Size	Download
1. clt_A1_1950_2000.nc		NetCDF	gridded	45145656	□ LLNL HRM
2. hur_A1_1950_2000.nc		NetCDF	gridded	437489896	□ LLNL HRM
3. hus_A1_1950_2000.nc		NetCDF	gridded	437489932	□ LLNL HRM
4. mrsos_A1_1950_2000.nc		NetCDF	gridded	45146084	□ LLNL HRM
5. pr_A1_1950_2000.nc		NetCDF	gridded	45145748	□ LLNL HRM
6. prc_A1_1950_2000.nc		NetCDF	gridded	45145744	□ LLNL HRM
7. prsn_A1_1950_2000.nc		NetCDF	gridded	45145656	□ LLNL HRM
8. ps_A1_1950_2000.nc		NetCDF	gridded	45145652	□ LLNL HRM
9. psl_A1_1950_2000.nc		NetCDF	gridded	45145664	□ LLNL HRM
10. snd_A1_1950_2000.nc		NetCDF	gridded	45145656	□ LLNL HRM
11. ta_A1_1950_2000.nc		NetCDF	gridded	437489884	□ LLNL HRM
12. tas_A1_1950_2000.nc		NetCDF	gridded	45145868	□ LLNL HRM
13. ts_A1_1950_2000.nc		NetCDF	gridded	45145660	□ LLNL HRM
14. ua_A1_1950_2000.nc		NetCDF	gridded	437489928	□ LLNL HRM
15. uas_A1_1950_2000.nc		NetCDF	gridded	45145872	□ LLNL HRM
16. va_A1_1950_2000.nc		NetCDF	gridded	437489936	□ LLNL HRM
17. vas_A1_1950_2000.nc		NetCDF	gridded	45145880	□ LLNL HRM
18. wap_A1_1950_2000.nc		NetCDF	gridded	437489912	□ LLNL HRM
19. zg_A1_1950_2000.nc		NetCDF	gridded	437489896	□ LLNL HRM

1-43 of 19 datafiles

Current ESG Sites



Primary ESG Servers

Mass storage, disk cache, and computation

Web and applications-based access to management, discovery, analysis, and visualization

PMEL: applications

NCAR: Climate change prediction and data archive

LBNL/NERSC: Climate data archive

LLNL: Model diagnostics and inter-comparison

USC/ISI: Globus, grid applications, and metadatabases

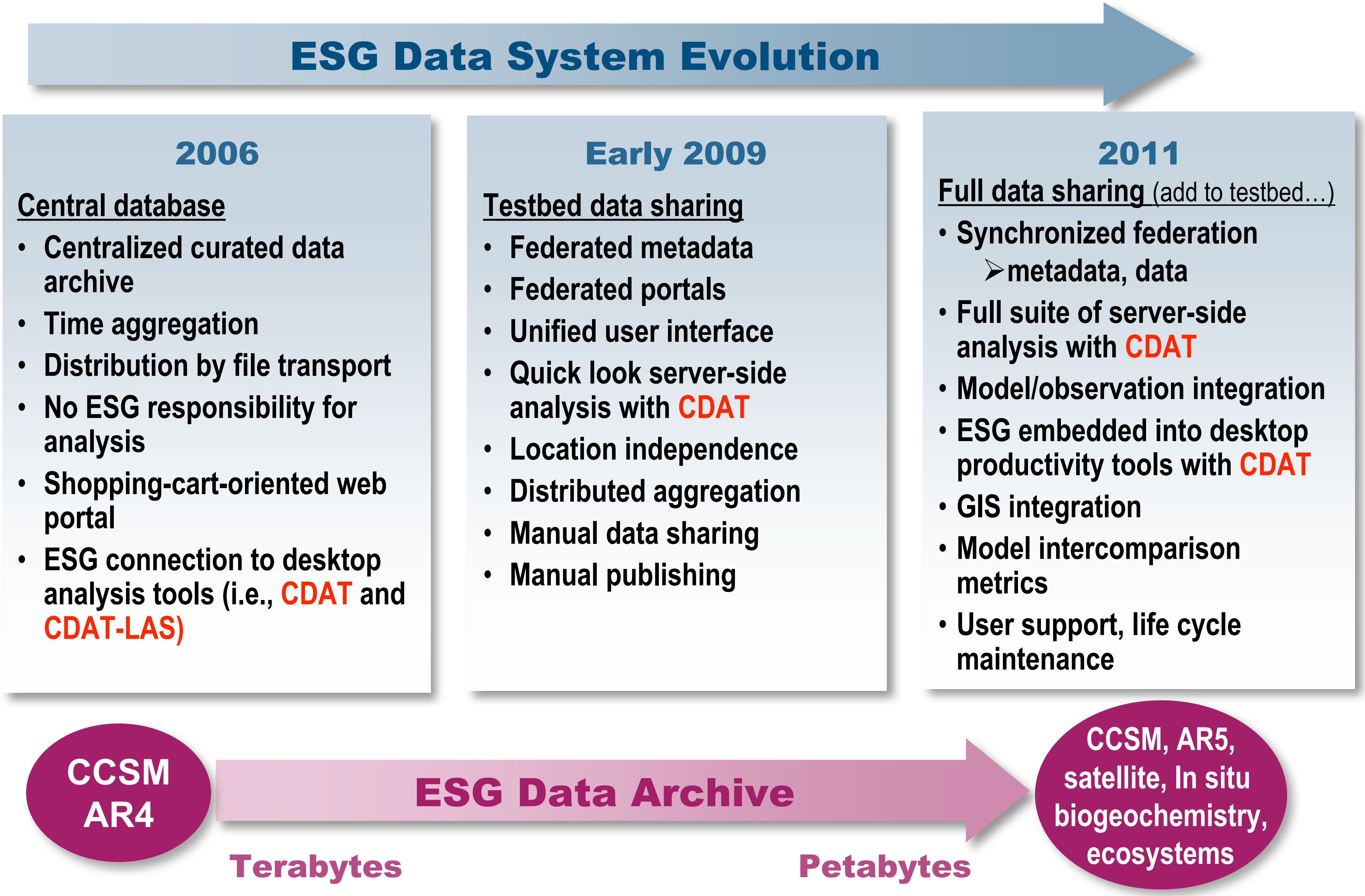
LANL: Climate and ocean data archive

ANL: Globus and grid applications

ORNL: Simulation and climate data archive

Evolving ESG for the future

Earth System Grid



ESG Data System Evolution

2006

<u>Central database</u>

- Centralized curated data archive
- Time aggregation
- Distribution by file transport
- No ESG responsibility for analysis
- Shopping-cart-oriented web portal
- ESG connection to desktop analysis tools (i.e., CDAT and CDAT-LAS)

Early 2009

<u>Testbed data sharing</u>

- Federated metadata
- Federated portals
- Unified user interface
- Quick look server-side analysis with CDAT
- Location independence
- Distributed aggregation
- Manual data sharing
- Manual publishing

2011

<u>Full data sharing (add to testbed…)</u>

- Synchronized federation
  - ➢metadata, data
- Full suite of server-side analysis with CDAT
- Model/observation integration
- ESG embedded into desktop productivity tools with CDAT
- GIS integration
- Model intercomparison metrics
- User support, life cycle maintenance

CCSM AR4

ESG Data Archive

CCSM, AR5, satellite, In situ biogeochemistry, ecosystems

Terabytes

Petabytes

The growing importance of climate simulation data standards

- Global Organization for Earth System Science Portal (GO-ESSP)
  - International collaboration to develop new generation of software infrastructure
  - Access to observed and simulated data from climate and weather communities
  - Working closely together using agreed upon standards
  - Last Annual meeting held at PCMDI
- NetCDF Climate and Forecast (CF) Metadata Convention standards
  - Specify syntax and vocabulary for climate and forecast metadata
  - Promotes the processing and sharing of data
  - The use of CF was essential for the success of the IPCC data dissemination

Supporting CF and CMOR

Earth System Grid

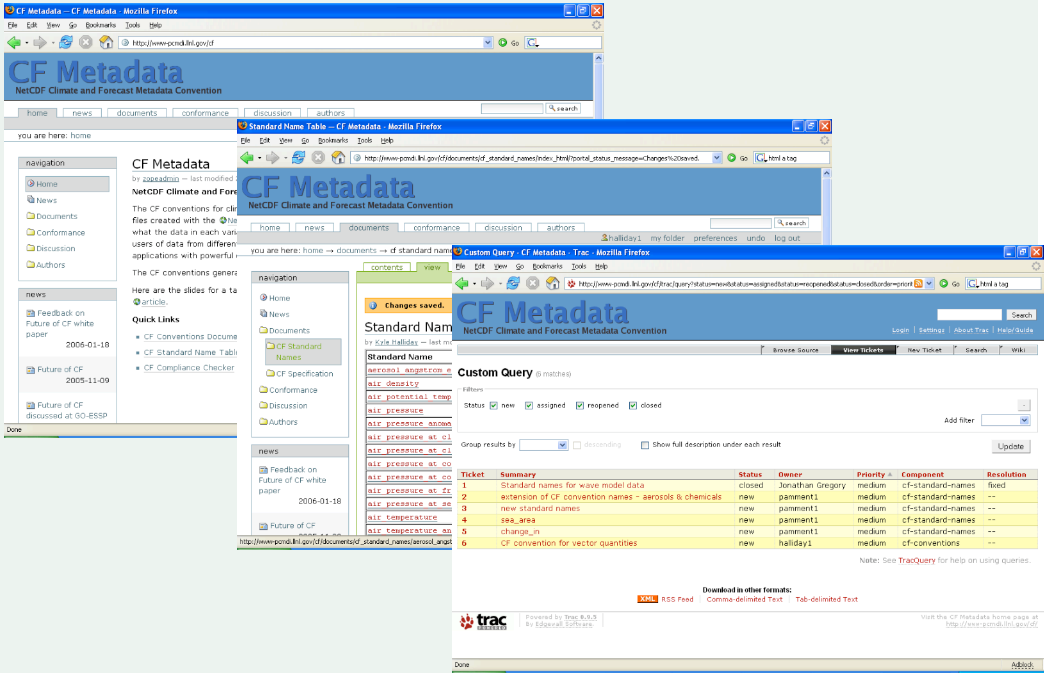
Future issues for CF

- Develop further fundamental tools (such as Climate Model Output Rewriter - CMOR)
- Develop staggered and unstructured grids
- Deliver netCDF data into Geographical Information Systems (GIS)
- Upgrade to netCDF-4
- Include in situ observations

CF/CMOR Development

- New CF website developed by PCMDI
- repository
  - ➢ News
  - ➢ Documents
    - ✓ CF Conventions
    - ✓ CF Standard Name table
- Conformance
  - ➢ Requirements & Recommendations
  - ➢ CF Compliance Checker
- Mailing List
  - ➢ Archives

New CF website



CF Metadata — CF Metadata - Mozilla Firefox
http://www-pcmdi.llnl.gov/cf
CF Metadata
NetCDF Climate and Forecast Metadata Convention
home
news
documents
conformance
discussion
authors
search
you are here: home
navigation
Home
News
Documents
Conformance
Discussion
Authors
news
Feedback on Future of CF white paper 2006-01-18
Future of CF 2005-11-09
Future of CF discussed at GO-ESSP
Done
Standard Name Table — CF Metadata - Mozilla Firefox
http://www-pcmdi.llnl.gov/cf/documents/cf_standard_names/index_html/?portal_status_message=Changes%20saved.
CF Standard Names
CF Specification
Custom Query - CF Metadata - Trac - Mozilla Firefox
Login | Settings | About Trac | Help/Guide
Browse Source
View Tickets
New Ticket
Search
Wiki

CF Metadata

by zopeadmin — last modified

NetCDF Climate and Fore

The CF conventions for cli files created with the Ne what the data in each vari users of data from differen applications with powerful

The CF conventions genera

Here are the slides for a ta article.

Quick Links

- CF Conventions Docume
- CF Standard Name Tabl
- CF Compliance Checker

Changes saved.

Standard Nam

by Kyle Halliday — last mo

Standard Name
aerosol_angstrom_e
air_density
air_potential_temp
air_pressure
air_pressure_anoma
air_pressure_at_cl
air_pressure_at_cl
air_pressure_at_co
air_pressure_at_co
air_pressure_at_fr
air_pressure_at_se
air_temperature
air_temperature_an

Custom Query (6 matches)

Filters

Status ☑ new ☑ assigned ☑ reopened ☑ closed

Add filter

Group results by □ descending □ Show full description under each result

Update

Ticket	Summary	Status	Owner	Priority ▲	Component	Resolution
1	Standard names for wave model data	closed	Jonathan Gregory	medium	cf-standard-names	fixed
2	extension of CF convention names - aerosols & chemicals	new	pamment1	medium	cf-standard-names	--
3	new standard names	new	pamment1	medium	cf-standard-names	--
4	sea_area	new	pamment1	medium	cf-standard-names	--
5	change_in	new	pamment1	medium	cf-standard-names	--
6	CF convention for vector quantities	new	halliday1	medium	cf-conventions	--

Note: See TracQuery for help on using queries.

Download in other formats:

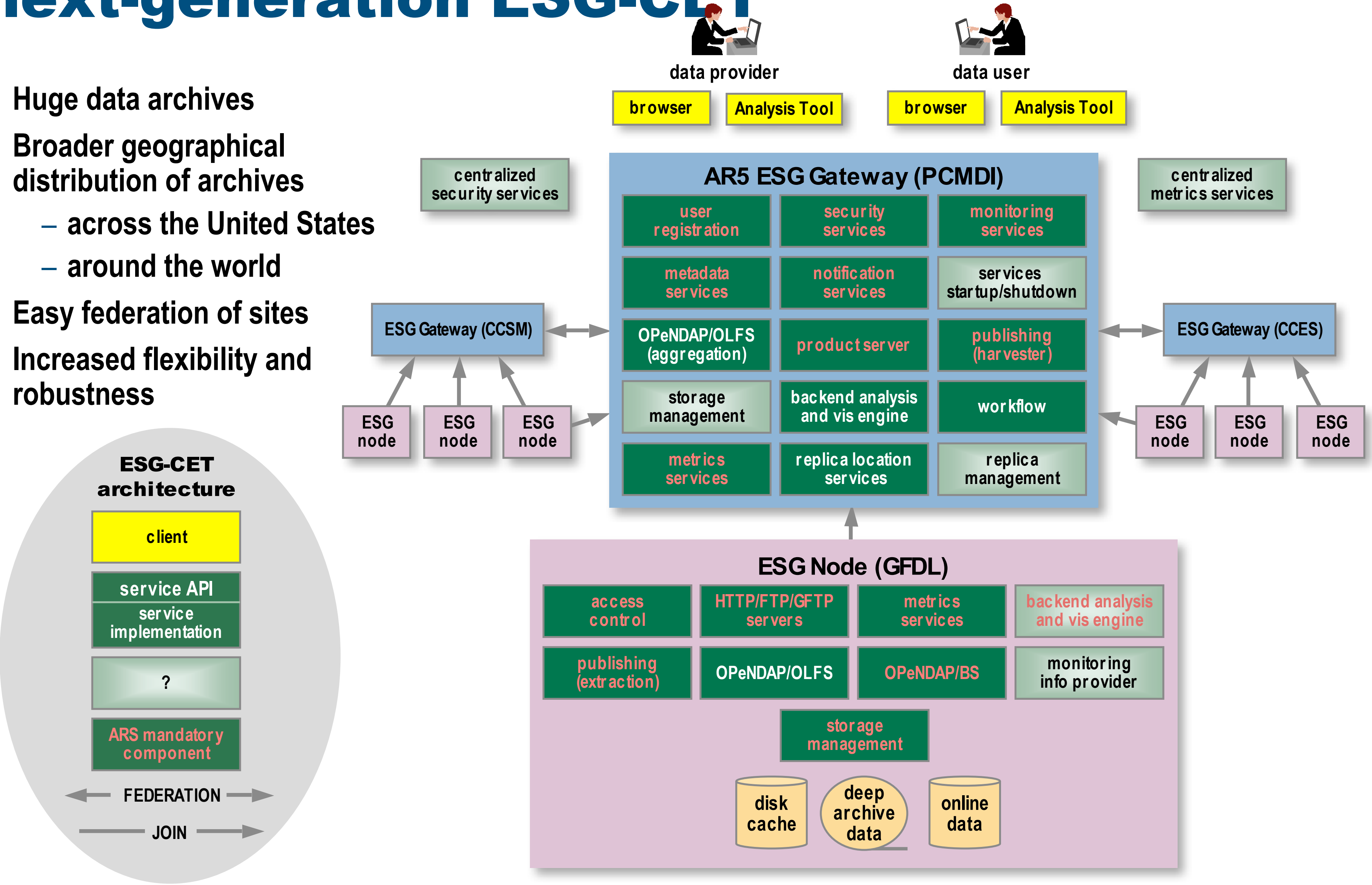
XML RSS Feed | Comma-delimited Text | Tab-delimited Text

Powered by Trac 0.9.5
By Edgewall Software.

Visit the CF Metadata home page at http://www-pcmdi.llnl.gov/cf/

Architecture of the next-generation ESG-CET

- Huge data archives
- Broader geographical distribution of archives
  - across the United States
  - around the world
- Easy federation of sites
- Increased flexibility and robustness



UCRL-PRES-226277

Climate Data Analysis Tools: Software for Distributed Model Diagnosis & Intercomparison Research

PCMDI Software Team

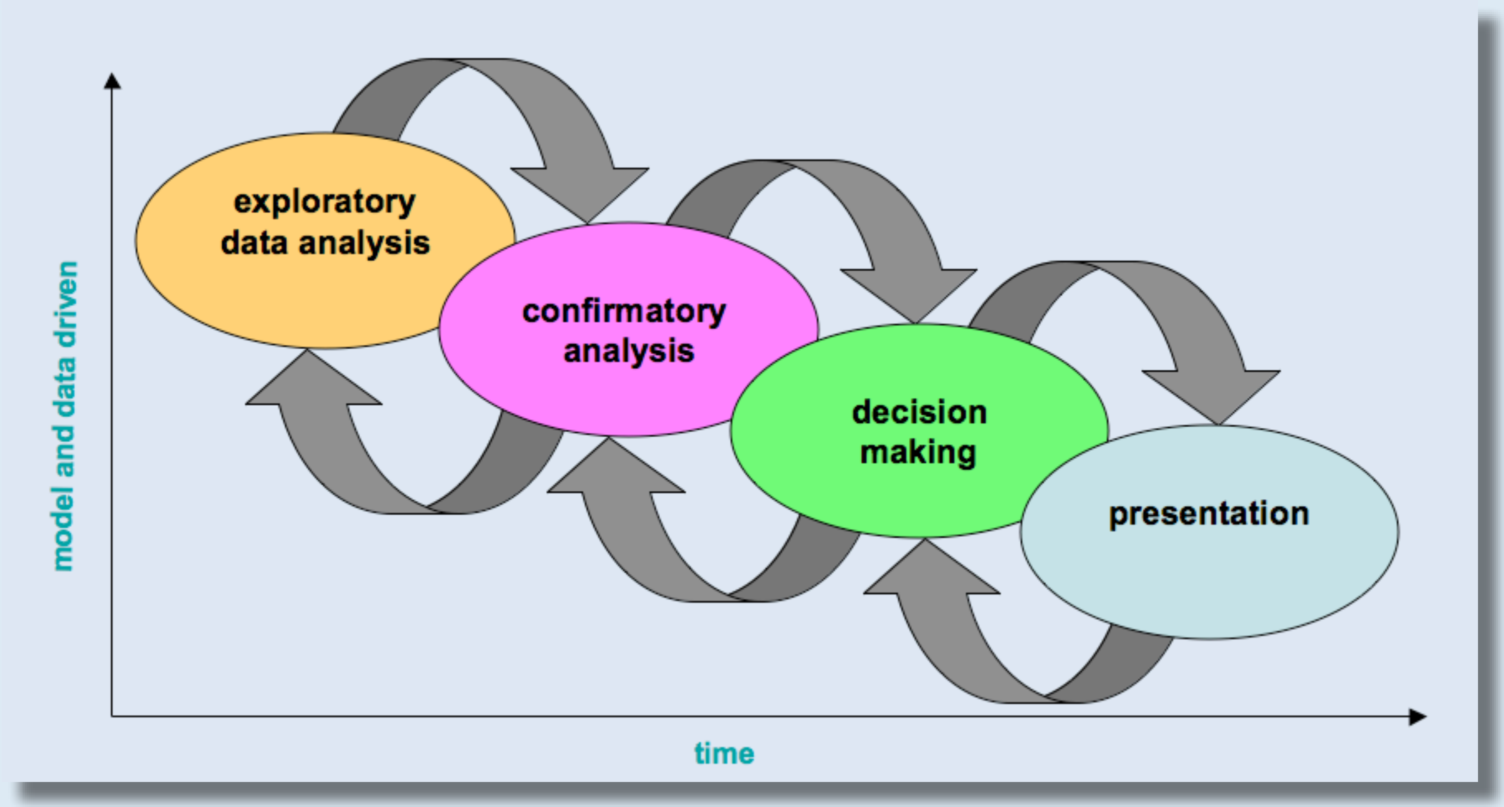
Challenges facing CDAT

- Integrating CDAT into a distributed environment
- Providing climate diagnostics
- Delivering climate component software to the community
- Working with other forms of climate Metadata describing extended modeling simulations (e.g., atmospheric aerosols and chemistry, carbon cycle, dynamic vegetation, etc.)
- Testbed needed by late 2008 – early 2009

CDAT objectives

CDAT Objectives

Seamless mechanisms for climate information exploration and analysis.



Enabling data management, data analysis, and visualization for intercomparison research

Climate Data Analysis Tools (CDAT)

CDAT Goal

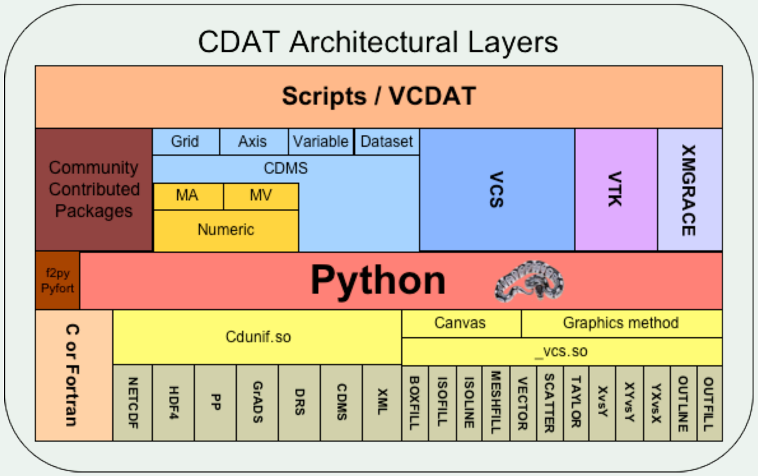
Address the challenges of enabling data management, discovery, access, and advanced data analysis for climate model diagnosis and intercomparison research.

Typical usage examples of CDAT

- Calculate a long-term average
- Define wind-speed from u- and v-components
- Subset a dataset, selecting a spatiotemporal region
- Aggregate 1000s of files into a small XML file
- Generate a Hovmoller plot

What is CDAT?

- CDAT *IS* Python!
- Designed for climate science data
- Scriptable
- Open-source and free



CDAT Architectural Layers

Scripts / VCDAT

Community Contributed Packages

Grid

Axis

Variable

Dataset

CDMS

MA

MV

Numeric

VCS

VTK

XMGRACE

f2py Pyfort

Python

C or Fortran

Cdunif.so

Canvas

Graphics method

_vcs.so

NETCDF

HDF4

PP

GrADS

DRS

CDMS

XML

BOXFILL

ISOFILL

ISOLINE

MESHFILL

VECTOR

SCATTER

TAYLOR

XvsY

XYvsY

YXvsX

OUTLINE

OUTFILL

Evolving CDAT into an integrated client technology workplace

Climate Data Analysis Tools (CDAT)

CDAT Integrated Analysis Evolution

2006

Community software

- Python based
- Start to finish environment
- Diverse analysis tools
- Languages: C/C++, Java, FORTRAN, Python
- Platforms: Unix, Mac, Windows
- VCDAT: discover, learn, and browse with a few clicks
- Connection to ESG

Early 2009

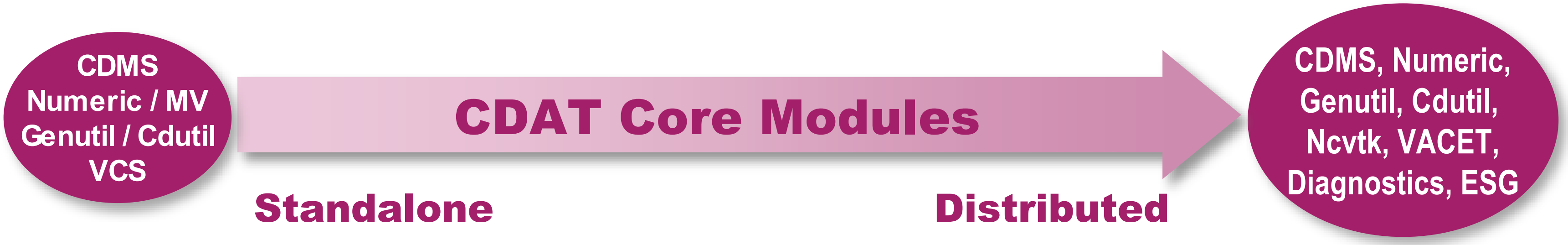
Testbed distributed analysis

- Equal-access to shared resources (Web/Grid services)
- Quick look server-side analysis tool for ESG
- Diagnostics specific to AR5
- GFDL Ncvtk 3D visualization
- Web-CDAT: discover, learn, and browse via web browser
- Serving Google Maps and Google Earth Data with CDAT

2011

Full analysis sharing

- Full suite server-side analysis tool for ESG
- ESG embedded into desktop productivity tools (i.e., CDAT)
- GIS integration with CDAT
- SciDAC VACET analysis and visualization collaboration
- Global Organization for Earth System Science Portal (GO-ESSP)
- Remote generic apps for ESG



CDMS
Numeric / MV
Genutil / Cdutil
VCS

CDAT Core Modules

CDMS, Numeric, Genutil, Cdutil, Ncvtk, VACET, Diagnostics, ESG

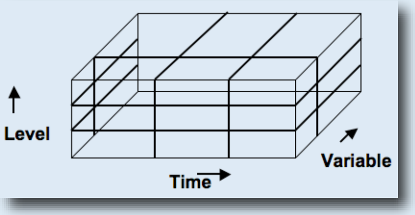
Standalone

Distributed

CDAT examples

CDSCAN

- Data aggregation: collections of files/datasets are treated as single entities.
- Aspects of aggregation:
  - combining/merging variables,
  - joining variables,
  - new coordinate axes,
  - overlaying/adding metadata,
  - nesting datasets



- PCMDI CDAT supports aggregations via the cdscan utility that uses XML representation
- cdscan will analyse the archive for:
  - variable information
  - axis information
  - global (universal) metadata
- Why use cdscan
  - Large datasets described as a grouped entity.
  - No need to know underlying data format.
  - No need to know file-names.
  - Datasets can be sliced in any way the user chooses using logical spatio-temporal selectors rather than loops of programming code.
  - You can use it to improve the metadata of your data files…
- cdscan in action
  - $ cdscan –x monthly_means.xml ./*.nc

MV

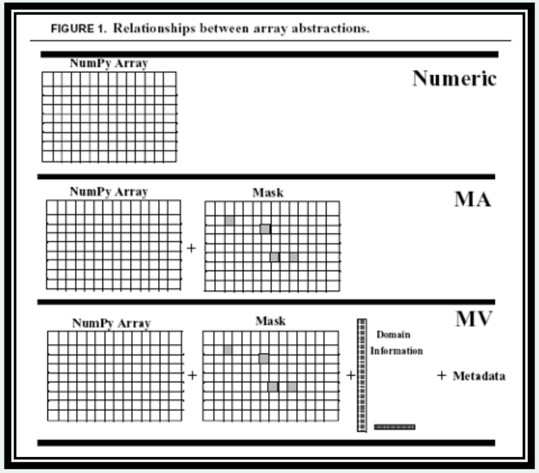


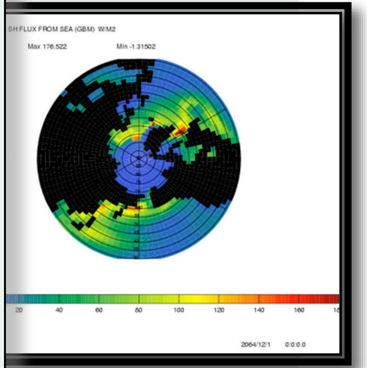
FIGURE 1. Relationships between array abstractions.

```
>>> import cdms, MV
>>> f_surface = cdms.open('sftlf_ta.nc')
>>> surf = f_surface('sftlf')

# Designate land where "surf" has values
# not equal to 100
>>> land_only = MV.masked_not_equal(surf, 100.)
>>> land_mask = MV.getmask(land_only)

# Now extract a variable from another file
>>> f = cdms.open('ta_1994-1998.nc')
>>> ta = f('ta')

# Apply this mask to retain only land values.
>>> ta_land = cdms.createVariable(ta,
        mask=land_mask, copy=0, id='ta_land')
```

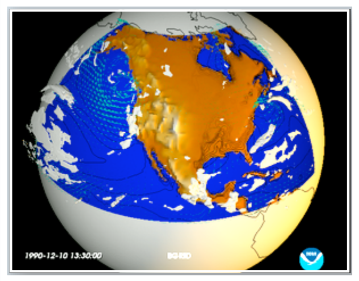


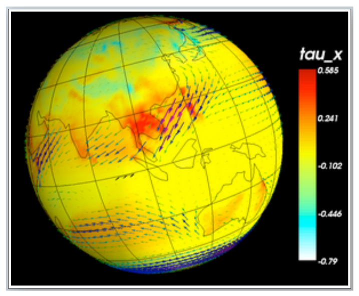
CDAT examples

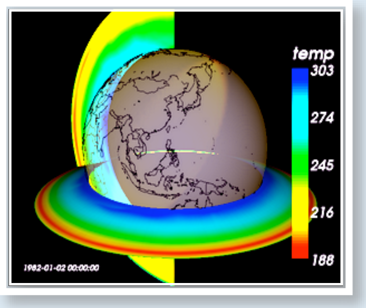
Ncvtk

Collaboration:

CDAT developers are currently working with Ncvtk developers to make Ncvtk 3D graphics accessible to the CDAT community. Ncvtk is a collection of commonly used 3D visualization methods applied to data on structured lat/lon grids.

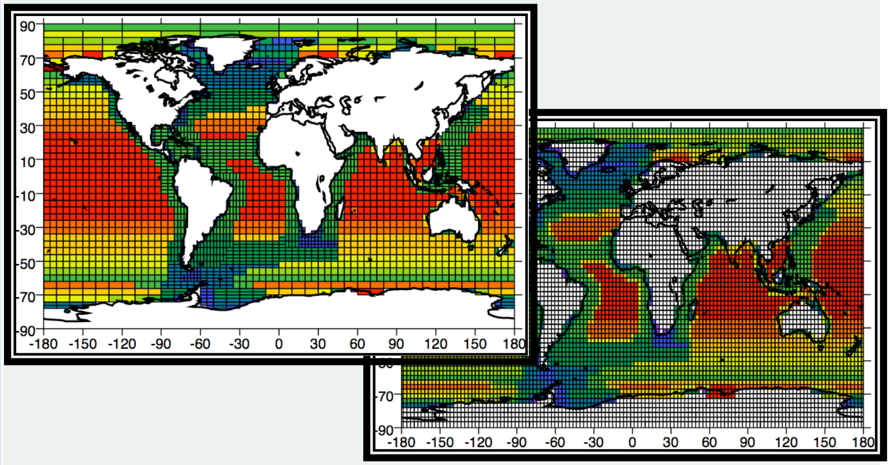






Regridder

```
#!/usr/local/cdat/bin/python
import cdms
from regrid import Regridder
f = cdms.open('temp.nc')
t= f.variables['t']
ingrid = t.getGrid()
outgrid = cdms.createUniformGrid( -90.0, 46, 4.0, 0.0, 72, 5.0)
regridFunc = Regridder(ingrid, outgrid)
newt = regridFunc(t)
import vcs
vcs.init().plot(t)
vcs.init().plot(newt)
```



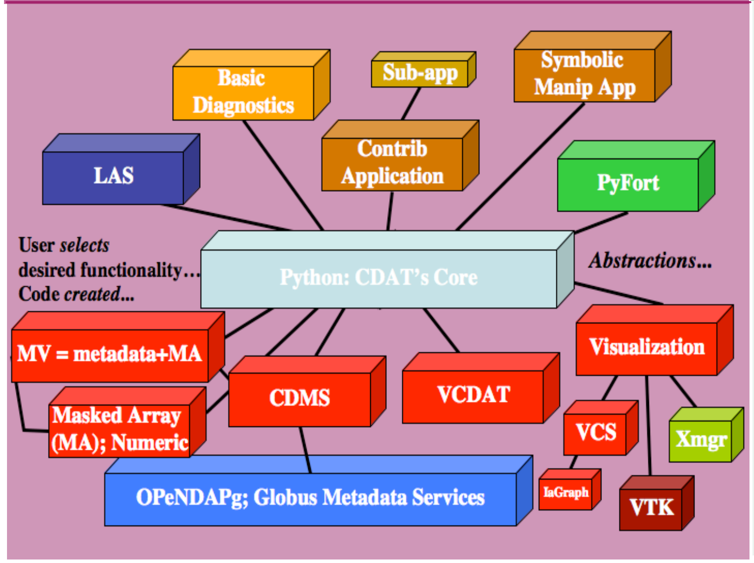
CDAT facts and figures

Climate Data Analysis Tools (CDAT)

CDAT Users

- Over 120 mailing list registers
  - ➤Probably 7 to 10 times more casual users
- Mailing list archive: over 1,000 message (~30 per month)
- 912 Downloads since May 19, 2006 for version 4.1
- Improved Documentation

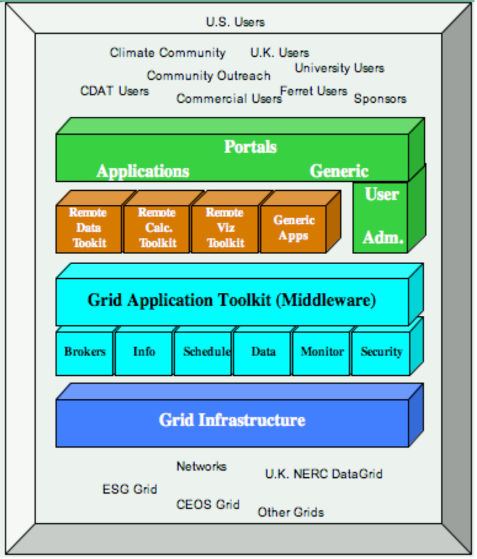
CDAT Core Modules



CDAT Collaborations

Some CDAT development centers:

- British Atmospheric Data Center
- LBNL
- GFDL
- Laboratory of Science of Climate and the Environment (LSCE), FR
- PCMDI
- University of Chicago
- University of Hawaii
- University of Reading, UK



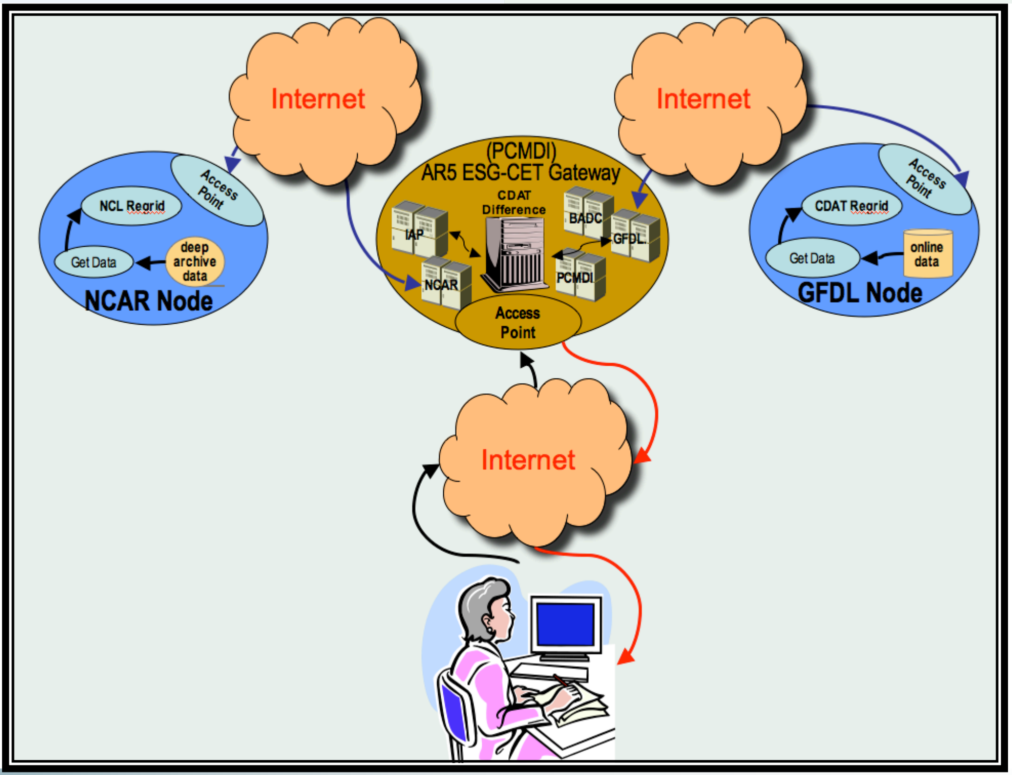
Simple intercomparison use case scenario

Current Scenario

- Browse PCMDI’s centralized database
- Download data
- Organize data on local site
- Regrid data at local site
- Perform diagnostics
- Produces results

Future Scenario

- Search, browse and discover distributed data
- Remote site
  - ➢Request data
  - ➢Regrids
  - ➢Diagnostics
- ESG returns results


Internet
Internet
Internet
Access Point
NCL Regrid
Get Data
deep archive data
NCAR Node
(PCMDI)
AR5 ESG-CET Gateway
CDAT Difference
IAP
BADC
GFDL
NCAR
PCMDI
Access Point
Access Point
CDAT Regrid
Get Data
online data
GFDL Node